Topical Stance Detection for Twitter: A Two-Phase LSTM Model Using Attention

Kuntal Dey, Ritvik Shrivastava, Saroj Kaushik

[1] Kuntal Dey, IBM Research, New Delhi, India. kuntadey@in.ibm.com
[2] Ritvik Shrivastava, Netaji Subhas Institute of Technology, New Delhi, India. ritviks.it@nsit.net.in
[3] Saroj Kaushik, Indian Institute of Technology, Delhi, India. saroj@cse.iitd.ac.in

Abstract. The topical stance detection problem addresses detecting the stance of the text content with respect to a given topic: whether the sentiment of the given text content is in FAVOR of (positive), is AGAINST (negative), or is NONE (neutral) towards the given topic. Using the concept of attention, we develop a two-phase solution. In the first phase, we classify subjectivity - whether a given tweet is neutral or subjective with respect to the given topic. In the second phase, we classify sentiment of the subjective tweets (ignoring the neutral tweets) - whether a given subjective tweet has a FAVOR or AGAINST stance towards the topic. We propose a Long Short-Term memory (LSTM) based deep neural network for each phase, and embed attention at each of the phases. On the SemEval 2016 stance detection Twitter task dataset [6], we obtain a best-case macro F-score of 68.84% and a best-case accuracy of 60.2%, outperforming the existing deep learning based solutions. Our framework, T-PAN, is the first in the topical stance detection literature, that uses deep learning within a two-phase architecture.

1 Introduction

Twitter, a hotbed of user generated content, has recently found traction among the researchers for the problem of topical stance detection. Topical stance detection is the problem of finding whether a given tweet takes a FAVOR (positive), AGAINST (negative) or NONE (neutral) stance towards a given topic. It is at core of the opinion polarity detection and mining problem. The problem is useful to solve in several practical scenarios, such as detecting user stance towards aspects of political, economic and social events, understanding stance-specific information propagation behavior of users *etc*.

1.1 Related Work

Sentiment detection from user-generated content has been a long-standing problem [9] [8]. However, stance detection, where the sentiment (opinion) of the user is not generic but with respect to a specific topic, has gained research attention only in recent times. A seminal work by Mohammed *et al.* [7], followed by a SemEval 2016 task [6] conducted by the authors, resulted in starting wide-spread research in the area.

Different models, including traditional machine learning approaches, genetic algorithms, and deep learning approaches such as convolutional neural networks (CNN),

recurrent neural networks (RNN) and long short-term memory (LSTM), were proposed in the SemEval 2016 topical stance detection contest. MITRE [13] provided the best deep learning solution in the contest, initializing weights from a 256-dimensional word embeddings learned learned using the word2vec skip-gram algorithm [5], followed by a second layer with 128 LSTM units. Among the other works, *pkudblab* [11] and DeepStance [10] use deep CNN models.

Some works used a two-phase approach. ECNU [14], in the first phase, determines whether a given tweet is relevant to a given target topic, and in the second phase, detects orientation (favor/against). The work by ltl.uni-due [12] also uses a two-level stacked classifier approach using Support Vector Machines (SVM). Among others, TakeLab [1], mixed machine learning with genetic algorithms. Other approaches, such as CU-GWU [3] and IUCL-RF [4], employed traditional machine learning.

The overall average values of F-scores, obtained by the task participants, ranged from 46.19 at the lower end to all the way up to 67.82 at the higher end. A recent work was conducted by Du *et al.* [2], the first of its kind that deeply ingrained the stance words in the architecture and used attention modeling. It outperformed the deep learning based approaches, attaining F-score of 68.79% as against the deep-learning state of the art F-score of 67.82%. We further observe that, the SemEval 2016 tasks were evaluated as a macro average of the F-score for only the *favor* and *against*, ignoring the *none* (neutral) class. We, however, perform accuracy measurements against all the three classes as well (in addition to the F-score that we measure following the traditional literature), and show that our model outperforms the best-known deep learning system not only for two-class macro average F-score, but for a full three-class accuracy measure as well.

1.2 Our Contributions

We propose a two-phase approach, using attention embedding at each phase and encoding using LSTM. The given SemEval 2016 [6] dataset contains three classes - FAVOR, AGAINST and NONE. Our work is based on the observation that messages with neutral stances are usually non-subjective, while the ones with favor and against stances are usually subjective. Thus, in the first phase of our two-phase approach, we use a LSTM to detect subjectivity, and classify into subjective (non-neutral) versus neutral (none). And in the second phase, we use another LSTM to detect sentiment (favor/against) of the tweets that were labeled subjective in the earlier phase. Akin to the philosophy of Du *et al.* [2], we also use an attention model, and deeply embed the topical attention as part of the input to the classifier. Since a given tweet does not necessarily contain the topic against which the stance is sought for, this step plays an important role in transforming the learning into a topic-specific learning. This is absent in the literature except for Du *et al.* [2]. Our model thus is the first of its kind, that uses a two-phase LSTM-based architecture with attention embedding ingrained.

The contributions of our work are the following.

- We propose T-PAN, a two-phase attention-embedded LSTM-based approach for detecting stance of tweets towards given topics.
- In the first phase, we perform subjectivity analysis of the tweets, using a combination of LSTM and attention embedding.

– In the second phase, we perform sentiment analysis on the subjective tweets, again using a combination of LSTM and attention embedding.
– Empirically, on the SemEval 2016 benchmark dataset, we demonstrate the effectiveness of our system. Our model is novel, and we outperform the deep learning based literature in terms of accuracy (ours is 60.2% while the deep learning literature is at 58.7%), as well as F-score (ours is 68.84% while the deep learning literature is at 68.79%).

2 Central Idea

2.1 Approach Overview

Table 1 shows a few randomly chosen samples from the training set across topics, to provide the reader with an intuition of the data available. As mentioned earlier, our task comprises of three classes of data: FAVOR, AGAINST and NONE. Since *favor* and *against* tweets are often subjective in nature, while the *neutral* tweets often are non-subjective. The architecture of our system is presented on Figure 1.

Our model is a two-phase one. At each phase, there are two components - a bi-directional LSTM and a attention logic. The bi-directional LSTM is used for feature encoding. The attention logic uses augmentation of the word embeddings with target topics, and subsequently passes it through a linear layer for computing attention of each word in the text in the context of the topic under consideration.

Target	**Tweet**	**Stance**
Examples from the **favor** *stance*		
Atheism	Everyone is able to believe in whatever they want. #Freedom	FAVOR
Feminist Movement	@OliviaJeniferx it's not always the guys job. #equality	FAVOR
Examples from the **against** *stance*		
Atheism	Be still. Be patient. Watch and let God work.	AGAINST
Feminist Movement	Friendly reminder that the "Gender Pay Gap" is a myth.	AGAINST
Examples from the **none** *stance*		
Atheism	Alot of angry people in this world. Peace to all. #love	NONE
Feminist Movement	@sass_unicorn lol! Young male children for	NONE

Table 1. Random examples of tweets of the different stances, for a few of the given target topics

2.2 Embedding Augmentation with Target Topics and Determining Attention

To compute attention, we augment the embedding of the constituent words with the average embedding of the target. If the words in a given target topic comprises of word embeddings $\{\tilde{z_1}, \tilde{z_2}, ..., \tilde{z_n}\}$, then we compute the embedding of the target topic $\tilde{z}$ as $\tilde{z} = \frac{\sum_{i=1}^{n} \tilde{z_i}}{|n|}$. The words within the sentence, that have the embeddings $\{z_1, z_2, ..., z_m\}$ of dimension d_z, are thus augmented with dimension $d_{\tilde{z}}$ (the dimension of $\tilde{z}$), and each word gets a new embedding dimension of $z + \tilde{z}$. This is processed as depicted in Figure 1, by first passing via a linear layer followed by a softmax, and subsequently ingraining the attention derived for each word into the LSTM-encoded features, using a product of the LSTM-embedded features and the output of the linear layer. We note that,

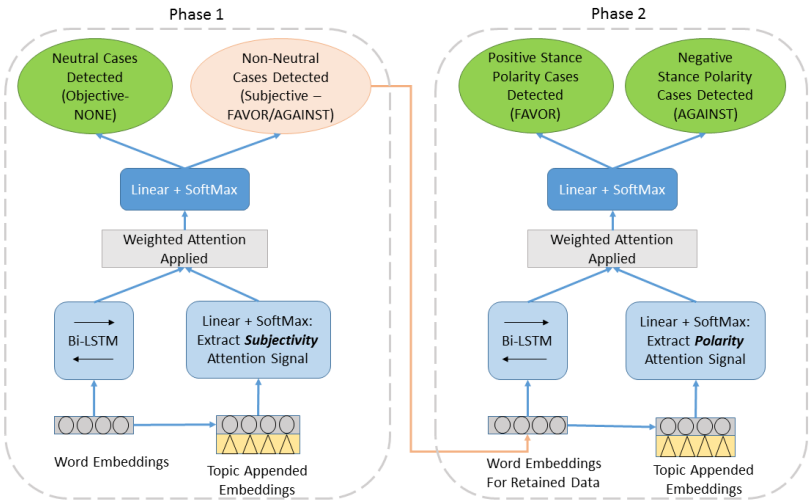

Fig. 1. System Architecture Diagram

while our approach is largely different from Du *et al.* [2] in terms of the overall system architecture (our approach is two-phase while their's is one-phase), the philosophy of augmenting each word of the sentence with the average embedding of the target topic words is similar.

2.3 Training the Models

Using a similar underlying architecture, the first phase is trained for subjectivity, and the second phase for sentiment polarity. Hence, the attention gets trained for subjectivity in the first phase and for polarity in the second phase. The subjective outputs of the first phase are passed through the second phase, while the rest (non-subjective) are assigned a class label NONE and kept aside. We try using both SGD (stochastic gradient descent) as well as Adam optimizers for experiments, and these yield similar effectiveness. We train our model using cross-entropy loss function. The loss of one phase is not propagated to the other.

3 Experiments

3.1 Data Description

We use the benchmark training and test data provided by the SemEval 2016 stance detection task [6]. For self-containment, we reproduce their data, in Table 2. We use the evaluation script they provide, for calculating F-score. Further, since their script only accounts for the FAVOR and AGAINST classes and computes a macro F-score as average of the two (ignoring the NONE class), we develop an additional script to calculate the accuracy using the three classes, as a ratio to the total number of correct predictions to the total test data size. We use PyTorch for programming. We perform data cleaning:

net slang removal (for tweet normalization) using an online dictionary[4] and stopword removal using a Stanford NLP resource for stopword removal[5].

			% of instances in Train				% of instances in Test		
Target	**#total**	**#train**	favor	against	neither	**#test**	favor	against	neither
Atheism	733	513	17.9	59.3	22.8	220	14.5	72.7	12.7
C.C.C.	564	395	53.7	3.8	42.5	169	72.8	6.5	20.7
Feminist Movement	949	664	31.6	49.4	19.0	285	20.4	64.2	15.4
Hillary Clinton	984	689	17.1	57.0	25.8	295	15.3	58.3	26.4
L.A.	933	653	18.5	54.4	27.1	280	16.4	67.5	16.1
All	4,163	2,914	25.8	47.9	26.3	1,249	24.3	57.3	18.4

Table 2. Data for the SemEval 2016 stance detection task. Target C.C.C. → Climate Change is Concern. Target L.A. → Legalization of Abortion. Table courtesy: [6].

3.2 Performance of Our Model T-PAN and Its Constituent Components

Our system delivers commendable performance for detecting the user stances towards the individual topics, as well as, a robust overall performance across the topics. We empirically observe the performance of our T-PAN model. We further examine the performance of different LSTM-based architectures that eventually are composed to develop our end-to-end framework. Table 3 provides the details of the performance attained by the full T-PAN model, as well as the impact of performance of the constituent LSTM blocks and configurations by systematic component ablation.

Phase 1	**Phase 2**	**Accuracy**
Bi-LSTM	Bi-LSTM	57.08
Bi-LSTM + Tweet Cleaning	Bi-LSTM + Tweet Cleaning	57.61
Bi-LSTM	One-Phase Attention	59.32
Bi-LSTM	One-Phase Attention + Tweet Cleaning	57.53
Bi-LSTM + Tweet Cleaning	One-Phase Attention + Tweet Cleaning	59.85
One-Phase Attention	One-Phase Attention	**60.22**
One-Phase Attention + Tweet Cleaning	One-Phase Attention + Tweet Cleaning	**60.24** (T-PAN)
Our implementation of TAN [2]		58.76

Table 3. Performance of the different underlying two-phase architectures.

3.3 Comparing Our System Against the Deep-Learning Literature

As observed in Table 4, our best system (the T-PAN model) outperforms the state of the art that uses deep neural networks for topical stance classification. Out of the five given classes, we perform the best in one class, the *TAN* model [2] outperforms us in two classes and the SemEval tasks perform better than our model (as well as better than the TAN model [2]) for the other two classes.

[4] http://www.noslang.com/dictionary

[5] https://nlp.stanford.edu/IR-book/html/htmledition/dropping-common-terms-stop-words-1.html

Target	NBOW	LSTM	LSTM$_E$	TOP Sem-Eval	TAN	*T-PAN*
Atheism	55.12	58.18	59.77	**61.47**	59.33	61.19
C.C.C.	39.93	40.05	48.98	41.63	53.59	**66.27**
Feminist Movement	50.21	49.06	52.04	**62.09**	55.77	58.45
Hillary Clinton	55.98	61.84	56.89	57.67	**65.38**	57.48
L.A.	55.07	51.03	60.34	57.28	**63.72**	60.21
Overall	60.19	63.21	66.24	67.82	68.79	**68.84**

Table 4. Comparing the different models. A part of the table has been replicated from Du *et al.* [2]. NBOW ← Neural Bag-of-Words. LSTM ← LSTM without target-specific embedding. LSTM$_E$ ← LSTM with target-specific embedding, by [2]. TOP Sem-Eval ← The best-reported systems in SemEval 2016. TAN ← The final output of [2]. T-PAN ← Our framework.

4 Conclusion

In this paper, we proposed T-PAN, a two-phase LSTM-based model with attention embedding, for detecting user stance with respect to given topics on Twitter. In the first phase, we classified the tweets into two: neutral and non-neutral, where the non-neutral stance comprised of favor (positive) and against (negative) stances. In the second, we classified the tweets labeled as non-neutral in the first phase, into two - favor and against stances. In each phase, we encoded the input sentences in form of a sequence of words using a bi-directional LSTM, and attention embedding. We further investigated the impact of embedding topical attention, as well as, the impact of different LSTM architectures, on the two-phase deep neural network based approach. We empirically demonstrated the robustness of our framework T-PAN, by delivering the highest-known performance among all the deep learning approaches present in the literature. Our model is easy to implement, reusable and practicable.

References

1. Boltuzic, F., Karan, M., Alagic, D., Šnajder, J.: Takelab at semeval-2016 task 6: Stance classification in tweets using a genetic algorithm based ensemble. SemEval pp. 464–468 (2016)
2. Du, J., Xu, R., He, Y., Gui, L.: Stance classification with target-specific neural attention networks. IJCAI pp. 3988–3994 (2017)
3. Elfardy, H., Diab, M.: Cu-gwu perspective at semeval-2016 task 6: Ideological stance detection in informal text. SemEval pp. 434–439 (2016)
4. Liu, C., Li, W., Demarest, B., Chen, Y., Couture, S., Dakota, D., Haduong, N., Kaufman, N., Lamont, A., Pancholi, M., et al.: Iucl at semeval-2016 task 6: An ensemble model for stance detection in twitter. SemEval pp. 394–400 (2016)
5. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
6. Mohammad, S.M., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C.: Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of SemEval. vol. 16 (2016)
7. Mohammad, S.M., Sobhani, P., Kiritchenko, S.: Stance and sentiment in tweets. arXiv preprint arXiv:1605.01655 (2016)
8. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S.M., Ritter, A., Stoyanov, V.: Semeval-2015 task 10: Sentiment analysis in twitter. Proceedings of SemEval-2015 (2015)
9. Rosenthal, S., Ritter, A., Nakov, P., Stoyanov, V.: Semeval-2014 task 9: Sentiment analysis in twitter. In: SemEval 2014. pp. 73–80 (2014)

10. Vijayaraghavan, P., Sysoev, I., Vosoughi, S., Roy, D.: Deepstance at semeval-2016 task 6: Detecting stance in tweets using character and word-level cnns. arXiv preprint arXiv:1606.05694 (2016)
11. Wei, W., Zhang, X., Liu, X., Chen, W., Wang, T.: pkudblab at semeval-2016 task 6: A specific convolutional neural network system for effective stance detection. SemEval pp. 384–388 (2016)
12. Wojatzki, M., Zesch, T.: ltl. uni-due at semeval-2016 task 6: Stance detection in social media using stacked classifiers. SemEval pp. 428–433 (2016)
13. Zarrella, G., Marsh, A.: Mitre at semeval-2016 task 6: Transfer learning for stance detection. arXiv preprint arXiv:1606.03784 (2016)
14. Zhang, Z., Lan, M.: Ecnu at semeval-2016 task 6: Relevant or not? supportive or not? a two-step learning system for automatic detecting stance in tweets. SemEval pp. 451–457 (2016)